Some remarks about metric spaces

Stephen William Semmes
Rice University
Houston, Texas

Contents

1 Basic notions and examples

Of course various kinds of metric spaces arise in various contexts and are viewed in various ways. In this brief survey we hope to give some modest indications of this. In particular, we shall try to describe some basic examples which can be of interest.

For the record, by a *metric space* we mean a nonempty set M together with a distance function $d(x, y)$, which is a real-valued function on $M \times M$ such that $d(x, y) \geq 0$ for all $x, y \in M$, $d(x, y) = 0$ if and only if $x = y$, $d(x, y) = d(y, x)$ for all $x, y \in M$, and

$$d(x, z) \leq d(x, y) + d(y, z) \tag{1.1}$$

for all $x, y, z \in M$. This last property is called the *triangle inequality*, and sometimes it is convenient to allow the weaker version

$$d(x, z) \leq C\,(d(x, y) + d(y, z)) \tag{1.2}$$

for a nonnegative real number C and all $x, y, z \in M$, which which case $(M, d(x, y))$ is called a quasi-metric space. Another variant is that we may wish to allow $d(x, y) = 0$ to hold sometimes without having $x = y$, in which case we have a semi-metric space, or a semi-quasi-metric space, as appropriate.

A sequence of points $\{x_j\}_{j=1}^\infty$ in a metric space M with metric $d(x, y)$ is said to converge to a point x in M if for every $\epsilon > 0$ there is a positive integer L such that

$$d(x_j, x) < \epsilon \quad \text{for all } j \ge L, \tag{1.3}$$

in which case we write

$$\lim_{j \to \infty} x_j = x. \tag{1.4}$$

A sequence $\{x_j\}_{j=1}^\infty$ of points in M is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there is a positive integer L such that

$$d(x_j, x_k) < \epsilon \quad \text{for all } j, k \ge L. \tag{1.5}$$

It is easy to see that every convergent sequence is a Cauchy sequence, and conversely a metric space in which every Cauchy sequence converges is said to be *complete*.

A very basic example of a metric space is the real line $\mathbf{R}$ with its standard metric. Recall that if x is a real number, then the *absolute value* of x is denoted $|x|$ and defined to be equal to x when $x \ge 0$ and to $-x$ when $x < 0$. One can check that

$$|x + y| \le |x| + |y| \tag{1.6}$$

and

$$|x\, y| = |x|\, |y| \tag{1.7}$$

when x, y are real numbers, and that the standard distance function $|x - y|$ on $\mathbf{R}$ is indeed a metric.

Let n be a positive integer, and let $\mathbf{R}^n$ denote the real vector space of n-tuples of real numbers. Thus elements x of $\mathbf{R}^n$ are of the form $(x_1, \ldots, x_n)$, where the n coordinates x_j, $1 \le j \le n$, are real numbers. If $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ are two elements of $\mathbf{R}^n$ and r is a real number, then the sum $x + y$ and scalar product $r\, x$ are defined coordinatewise in the usual manner, by

$$x + y = (x_1 + y_1, \ldots, x_n + y_n) \tag{1.8}$$

and

$$r\,x = (r\,x_1, \ldots, r\,x_n). \tag{1.9}$$

If x is an element of $\mathbf{R}^n$, then the standard Euclidean norm of x is denoted $|x|$ and defined by

$$|x| = \Big(\sum_{j=1}^n x_j^2\Big)^{1/2}. \tag{1.10}$$

One can show that

$$|x+y| \le |x| + |y| \tag{1.11}$$

holds for all $x, y \in \mathbf{R}^n$, and we shall come back to this in a moment, and clearly we also have that

$$|r\,x| = |r|\,|x| \tag{1.12}$$

for all $r \in \mathbf{R}$ and $x \in \mathbf{R}^n$, which is to say that the norm of a scalar product of a real number and an element of $\mathbf{R}^n$ is equal to the product of the absolute value of the real number and the norm of the element of $\mathbf{R}^n$. Using these properties, one can check that the standard Euclidean distance $|x-y|$ on $\mathbf{R}^n$ is indeed a metric.

More generally, a *norm* on $\mathbf{R}^n$ is a real-valued function $N(x)$ such that $N(x) \ge 0$ for all $x \in \mathbf{R}^n$, $N(x) = 0$ if and only if $x = 0$,

$$N(r\,x) = |r|\,N(x) \tag{1.13}$$

for all $r \in \mathbf{R}$ and $x \in \mathbf{R}^n$, and

$$N(x+y) \le N(x) + N(y) \tag{1.14}$$

for all $x, y \in \mathbf{R}^n$. If $N(x)$ is a norm on $\mathbf{R}^n$, then

$$d(x,y) = N(x-y) \tag{1.15}$$

defines a metric on $\mathbf{R}^n$. As for metrics, one can weaken the triangle inequality or relax the condition that $N(x) = 0$ implies $x = 0$ to get quasi-norms, semi-norms, and semi-quasi-norms.

Recall that a subset E of $\mathbf{R}^n$ is said to be *convex* if

$$t\,x + (1-t)\,y \in E \tag{1.16}$$

whenever x, y are elements of E and t is a real number such that $0 < t < 1$. A real-valued function $f(x)$ on $\mathbf{R}^n$ is said to be convex if and only if

$$f(t\,x + (1-t)\,y) \le t\,f(x) + (1-t)\,f(y) \tag{1.17}$$

for all $x, y \in \mathbf{R}^n$ and $t \in \mathbf{R}^n$ with $0 < t < 1$. If $N(x)$ is a real-valued function on $\mathbf{R}^n$ which is assumed to satisfy the conditions of a norm except for the triangle inequality, then one can check that the triangle inequality, the convexity of the closed unit ball

$$\{x \in \mathbf{R}^n : N(x) \leq 1\}, \tag{1.18}$$

and the convexity of $N(x)$ as a function on $\mathbf{R}^n$, are all equivalent.

For example, if p is a real number such that $1 \leq p < \infty$, then define $|x|_p$ for $x \in \mathbf{R}^n$ by

$$|x|_p = \Big(\sum_{j=1}^{n} |x|^p\Big)^{1/p}, \tag{1.19}$$

which is the same as the standard norm $|x|$ when $p = 2$. For $p = \infty$ let us set

$$|x|_\infty = \max\{|x_j| : 1 \leq j \leq n\}. \tag{1.20}$$

One can check that these define norms on $\mathbf{R}^n$, using the convexity of the function $|r|^p$ on $\mathbf{R}$ when $1 < p < \infty$ to check that the closed unit ball of $|x|_p$ is convex and hence that the triangle inequality holds when $1 < p < \infty$.

Let us now consider a class of metric spaces along the lines of Cantor sets. For this we assume that we are given a sequence $\{F_j\}_{j=1}^\infty$ of nonempty finite sets. We also assume that $\{\rho_j\}_{j=1}^\infty$ is a monotone decreasing sequence of positive real numbers which converges to 0.

For our space M we take the Cartesian product of the F_j's, so that en element x of M is a sequence $\{x_j\}_{j=1}^\infty$ such that $x_j \in F_j$ for all j. We define a distance function $d(x, y)$ on M by setting $d(x, y) = 0$ when $x = y$, and

$$d(x, y) = \rho_j \tag{1.21}$$

when $x_j \neq y_j$ and $x_i = y_i$ for all $i < j$. One can check that this does indeed define a metric space, and in fact $d(x, y)$ is an *ultrametric*, which is to say that

$$d(x, z) \leq \max(d(x, y), d(y, z)) \tag{1.22}$$

for all $x, y, z \in M$.

The classical Cantor set is the subset of the unit interval $[0, 1]$ in the real line obtained by first removing the open subinterval $(1/3, 2/3)$, then removing the the open middle thirds of the two closed intervals which remain, and so on. Alternatively, the classical Cantor set can be described as the

set of real numbers t such that $0 \le t \le 1$ and t has ab expansion base 3 whose coefficients are are either 0 or 2. This set, equipped with the standard Euclidean metric, is very similar to the general situation just described with each F_j having two elements and with $\rho_j = 2^{-j}$ for all j, although the metrics are not quite the same.

In general, if $(M, d(x,y))$ is a metric space and E is a nonempty subset of M, then E can be considered as a metric space in its own right, using the restriction of the metric $d(x,y)$ from M to E. Sometimes there may be another metric on E which is similar to the one inherited from the larger space M, and which has other nice properties, as in the case of Cantor sets just described. Another basic instance of this occurs with arcs in Euclidean spaces which are "snowflakes", and which are similar to taking the unit interval $[0,1]$ in the real line with the metric $|x-y|^a$ for some real number a, $0 < a < 1$, or other functions of the standard distance on $[0,1]$.

A nonempty subset E of a metric space $(M, d(x,y))$ is said to be *bounded* if the real numbers $d(x,y)$, $x, y \in E$, are bounded, in which case the *diameter* of E is denoted $\operatorname{diam} E$ and defined by

$$\operatorname{diam} E = \sup\{d(x,y) : x, y \in E\}. \tag{1.23}$$

A stronger condition is that E be *totally bounded*, which means that for each $\epsilon > 0$ there is a finite collection $A_1, \ldots, A_k$ of subsets of E such that

$$E \subseteq \bigcup_{j=1}^{k} A_j \tag{1.24}$$

and

$$d(x,y) < \epsilon \quad x, y \in A_j, \tag{1.25}$$

$j = 1, \ldots, k$. A basic feature of Euclidean spaces is that bounded subsets are totally bounded, and the generalized Cantor sets described before are totally bounded.

A metric space $(M, d(x,y))$ is *compact* if it is complete, so that every Cauchy sequence converges, and totally bounded. This is equivalent to the standard definitions in terms of open coverings or the existence of limit points. A closed and bounded subset of $\mathbf{R}^n$ is compact, and the generalized Cantor sets described earlier are compact.

Another way that metric spaces arise is to start with a connected smooth n-dimensional manifold M, which is basically a space which looks locally

like n-dimensional Euclidean space. At each point p in M one has an n-dimensional tangent space $T_p(M)$, which looks like $\mathbf{R}^n$ as a vector space, and on which one can put a norm. If at each point p in M one can identify $T_p(M)$ with $\mathbf{R}^n$ with its standard norm, then the space is Riemannian, and with general norms the space is of Riemann–Finsler type.

In this type of situation, the length of a nice path in M can be defined by integrating the infinitesimal lengths determined by the norms on the tangent spaces. The distance between two points is defined to be the infimum of the lengths of the paths connecting the two points. It is easy to see that this does indeed define a metric, with the triangle inequality being a consequence of the way that the distance is defined.

A basic example of this is the n-dimensional sphere $\mathbf{S}^n$ in $\mathbf{R}^{n+1}$, defined by

$$\mathbf{S}^n = \{x \in \mathbf{R}^{n+1} : |x| = 1\}. \tag{1.26}$$

The tangent space of $\mathbf{S}^n$ can be identified with the n-dimensional linear subspace of $\mathbf{R}^{n+1}$ of vectors x which are orthogonal to p as a vector itself. This leads to a Euclidean norm on the tangent spaces, inherited from the one on $\mathbf{R}^{n+1}$.

One can also define distances through paths in other situations. In a nonempty connected graph, every pair of points is connected by a path, the length of a path can be defined as the number of edges traversed, and the distance between two points can be defined as the length of the shortest path between the two points. In many standard fractals, like the Sierpinski gasket and carpet, there are a lot of paths of finite length between arbitrary elements of the fractal, and the infimum of the lengths of these paths defines a metric on the fractal.

Let $(M, d(x, y))$ be a metric space. If A, B are nonempty subsets of M and t is a positive real number, then let us say that A, B are "t-close" if for every $x \in A$ there is a $y \in B$ such that $d(x, y) < t$, and if for every $y \in B$ there is an $x \in A$ such that $d(x, y) < t$. By definition, this relation is symmetric in A and B.

A subset E of M is said to be *closed* if for every sequence $\{z_j\}_{j=1}^{\infty}$ of points in E which converges to a point z in M, we have that $z \in E$. Let us write $\mathcal{S}(M)$ for the set of nonempty closed and bounded subsets of M. If A, B are two elements of $\mathcal{S}(M)$, then the *Hausdorff distance* between A and B is defined to be the infimum of the set of positive real numbers t such that A, B are t-close.

Because A and B are bounded subsets of M, there are positive real numbers t such that A, B are t-close. The restriction to closed sets ensures that if A, B are t-close for all $t > 0$, then $A = B$. If A_1, A_2, A_3 are nonempty subsets of M and t_1, t_2 are positive real numbers such that A_1, A_2 are t_1-close and A_2, A_3 are t_2-close, then one can check that A_1, A_3 are $(t_1 + t_2)$-close, and this implies the triangle inequality for the Hausdorff distance.

From this it follows easily that $\mathcal{S}(M)$, equipped with the Hausdorff distance $D(A, B)$, is indeed a metric space. A basic result states that if M is compact, then $\mathcal{S}(M)$ is compact too. This is not too difficult to show.

Now let us consider a situation with a lot of deep mathematical structure which has been much-studied, involving interplay between algebra, analysis, and geometry.

Fix a positive integer n, which we may as well take to be at least 2. Let $\mathcal{M}_n^+$ denote the set of $n \times n$ real symmetric matrices which are positive-definite and have determinant equal to 1. One can think of this as a smooth hypersurface in the vector space of $n \times n$ real symmetric matrices, and it is a smooth manifold in particular.

We can start with a Riemannian view of this space. For each $H \in \mathcal{M}_n^+$, we can identify the tangent space of $\mathcal{M}_n^+$ at H with the vector space of $n \times n$ real symmetric matrices A such that the trace of $H^{-1} A$ is equal to 0. Thus for each such A, we get a one-parameter family of perturbations of H by taking $H + t\,A$, where t is a real number with small absolute value, so that $H + t\,A$ is still positive definite, and to first order in t these perturbations also have determinant equal to 1.

Conversely, to first order in t, each smooth perturbation of H in $\mathcal{M}_n^+$ is of this form. Now, for each A of this type, we define its norm as an element of the tangent space to $\mathcal{M}_n^+$ to be

$$\left(\operatorname{tr} H^{-1} A\, H^{-1} A\right)^{1/2}. \tag{1.27}$$

Here $\operatorname{tr} B$ denotes the trace of a square matrix A, and we are using ordinary matrix multiplication in this expression.

When H is the identity matrix, this reduces to the square root of $\operatorname{tr} A^2$, which is a kind of Euclidean norm of A. For general H's, we adapt the norm to H. It is still basically a Euclidean norm, so that we are in the Riemannian case.

Let us consider the transformation on $\mathcal{M}_n^+$ defined by $H \mapsto H^{-1}$. If $H + t\,A$ is a basic first-order deformation of H, as before, then that is transformed

to $(H + t\,A)^{-1}$, which is the same as

$$H^{-1} - t\,H^{-1}\,A\,H^{-1} \tag{1.28}$$

to first order in t. Thus A as a tangent vector at H corresponds to $-H^{-1}\,A\,H^{-1}$ as a tangent vector at H^{-1} under the mapping $H \mapsto H^{-1}$, and it is easy to see that the norm of A as a tangent vector at H is equal to the norm of $-H^{-1}\,A\,H^{-1}$ as a tangent vector at H^{-1}.

Now let T be an $n \times n$ matrix with determinant equal to 1, and let T^* denote its transpose, which is also an invertible $n \times n$ matrix. Associated to T is the mapping

$$H \mapsto T\,H\,T^*, \tag{1.29}$$

and if A corresponds to a tangent vector at H as before, then $T\,A\,T^*$ is the tangent vector at $T\,H\,T^*$ induced by our mapping. Again one can check that the norm of A as a tangent vector at H is the same as the norm of $T\,A\,T^*$ as a tangent vector at $T\,H\,T^*$.

Assume further that the entries of T are integers. This implies that the inverse of T also has integer entries, by Cramer's rule. The product of two such matrices has the same property, and indeed this defines a nice discrete group of matrices.

This discrete group acts on $\mathcal{M}_n^+$, and we can pass to the corresponding quotient space. That is, we now identify two elements H_1, H_2 in $\mathcal{M}_n^+$ if there is an integer matrix T as above so that

$$H_2 = T\,H_1\,T^*. \tag{1.30}$$

This relation between $H_1, H_2 \in \mathcal{M}_n^+$ is indeed an equivalence relation, so that we get a nice quotient.

Because of the discreteness of the group of integer matrices with determinant 1, the quotient space is still a smooth manifold, since it looks locally like $\mathcal{M}_n^+$. The norm on the tangent spaces still makes sense as well, because the transformations defining the equivalence relation preserves these norms, as we have seen. Thus this quotient of $\mathcal{M}_n^+$ is still a nice smooth connected Riemannian manifold.

In some situations like this the quotient space turns out to be compact. In this case the quotient space is not compact, but it does have finite volume. More precisely, the notion of volume at the level of the tangent spaces is determined by the norm, and preserved in the present circumstances by

the transformations used in the equivalence relation, and the volume of the quotient can be obtained by integrating the infinitesimal volumes.

The group of $n \times n$ matrices with integer entries and determinant 1 is a very interesting special case of discrete groups more generally. Suppose now that Γ is a group and that A is a finite symmetric subset of Γ which generates Γ, so that $\alpha^{-1} \in A$ when $\alpha \in A$ and every element of Γ can be expressed as a finite product of elements of A, with the identity element automatically corresponding to an empty product. This leads to the associated *Cayley graph*, in which two elements γ_1, γ_2 of Γ are considered to be adjacent if γ_2 can be expressed as $\gamma_1 \, \alpha$ for some $\alpha \in A$, and to a distance function on Γ which is invariant under left-multiplication in the group.

2 Aspects of analysis

Let $(M, d(x,y))$ be a metric space, let $f(x)$ be a real-valued function on M, and let C be a nonnegative real number. We say that f is *C-Lipschitz* if

$$|f(x) - f(y)| \le C \, d(x,y) \tag{2.1}$$

for all $x, y \in M$. This is equivalent to saying that

$$f(x) \le f(y) + C \, d(x,y) \tag{2.2}$$

for all $x, y \in M$. Notice that a function is 0-Lipschitz if and only if it is constant.

For instance, if p is a point in M, then $f_p(x) = d(x,p)$ is 1-Lipschitz, because

$$d(x,p) \le d(x,y) + d(y,p) \tag{2.3}$$

by the triangle inequality. More generally, if A is a nonempty subset of M, and if x is a point in M, then the distance from x to A is denoted $\mathrm{dist}(x,A)$ and defined by

$$\mathrm{dist}(x,A) = \inf\{d(x,y) : y \in A\}, \tag{2.4}$$

and one can check that $\mathrm{dist}(x,A)$ is a 1-Lipschitz function on M. If f_1, f_2 are two real-valued functions on M which are C_1, C_2-Lipschitz, respectively, and if α_1, α_2 are real numbers, then $\max(f_1, f_2)$, $\min(f_1, f_2)$ are C-Lipschitz with $C = \max(C_1, C_2)$, and $\alpha_1 \, f_1 + \alpha_2 \, f_2$ is C-Lipschitz with $C = |\alpha_1| \, C_1 + |\alpha_2| \, C_2$.

Now let C be a nonnegative real number and let s be a positive real number. A real-valued function $f(x)$ on M is said to be *C-Lipschitz of order s* if

$$|f(x) - f(y)| \le C\, d(x,y)^s \tag{2.5}$$

for all $x, y \in M$, which is again equivalent to

$$f(x) \le f(y) + C\, d(x,y)^s \tag{2.6}$$

for all $x, y \in M$. As before, $f(x)$ is 0-Lipschitz of order s if and only if $f(x)$ is constant on M.

When $0 < s < 1$, one can check that $d(x,y)^s$ is also a metric on M, which defines the same topology on M in fact. The main point in this regard is that the triangle inequality continues to hold, which follows from the observation that

$$(\alpha + \beta)^s \le \alpha^s + \beta^s \tag{2.7}$$

for all nonnegative real numbers α, β. A real-valued function $f(x)$ on M is C-Lipschitz of order s with respect to the metric $d(x,y)$ if and only if $f(x)$ is C-Lipschitz of order 1 with respect to $d(x,y)^s$, and as a result when $0 < s < 1$ one has the same statements for Lipschitz functions of order s as for ordinary Lipschitz functions.

When $s > 1$ the triangle inequality for $d(x,y)^s$ does not work in general, but we do have that

$$d(x,z)^s \le 2^{s-1}\,(d(x,y)^s + d(y,z)^s) \tag{2.8}$$

for all $x, y, z \in M$, because

$$(\alpha + \beta)^s \le 2^{s-1}\,(\alpha^s + \beta^s) \tag{2.9}$$

for all nonnegative real numbers α, β. Some of the usual properties of Lipschitz functions carry over to Lipschitz functions of order s, perhaps with appropriate modification, but it may be that the only Lipschitz functions of order s when $s > 1$ are constant. This is the case on Euclidean spaces with their standard metrics, and indeed a Lipschitz function of order $s > 1$ has first derivatives equal to 0 everywhere.

In harmonic analysis one considers a variety of classes of functions with different kinds of restrictions on size, oscillations, regularity, and so on, and these Lipschitz classes are fundamental examples. In particular, it can be

quite useful to have the parameter s available to adjust to the given circumstances. There are also other ways of introducing parameters to get interesting classes of functions and measurements of their behavior.

If M is the usual n-dimensional Euclidean space $\mathbf{R}^n$, with its standard metric, then one has the extra structure of translations, rotations, and dilations. If $f(x)$ is a real-valued function on $\mathbf{R}^n$ which is C-Lipschitz of order s, $f(x-u)$ is also C-Lipschitz of order s for each $u \in \mathbf{R}^n$, $f(\Theta(x))$ is C-Lipschitz of order s for each rotation Θ on $\mathbf{R}^n$, and $f(r^{-1}x)$ is $(C\,r^s)$-Lipschitz of order s for each $r > 0$. In effect, on general metric spaces we can consider classes of functions and measurements of their behavior which have analogous features, even if there are not exactly translations, rotations, and dilations.

A basic notion is to consider various scales and locations somewhat independently. In this regard, if $f(x)$ is a real-valued function on M, x is an element of M, and t is a positive real number, put

$$\operatorname{osc}(x,t) = \sup\{|f(y) - f(x)| : y \in M, d(y,x) \le t\}. \tag{2.10}$$

We implicitly assume here that $f(y)$ remains bounded on bounded subsets of M, so that this quantity is finite. For instance, f is C-Lipschitz of order s if and only if

$$t^{-s}\operatorname{osc}(x,t) \le C \tag{2.11}$$

for all $x \in M$ and $t > 0$.

Let us pause a moment and notice that

$$\operatorname{osc}(w,r) \le \operatorname{osc}(x,t) \tag{2.12}$$

when $d(w,x) + r \le t$. Thus,

$$r^{-s}\operatorname{osc}(w,r) \le 2^s\,t^{-s}\operatorname{osc}(x,t) \tag{2.13}$$

when $d(x,w) + r \le t$ and $r \ge t/2$. This is a kind of "robustness" property of these measurements of local oscillation of a function f on M. In particular, to sample the behavior of f at essentially all locations and scales, it is practically enough to look at a reasonably-nice and discrete family of locations and scales. For example, one might restrict one's attention to radii t which are integer powers of 2, and for a specific choice of t use a collection of points in M which cover suitably the various locations at that scale.

Instead of simply taking a supremum of some measurements of local oscillation like this, one can consider various sums of discrete samples of this

sort. This leads to a number of classes of functions and measurements of their behavior. One can adjust this further by taking into account the relation of some location and scale to some kind of boundaries, or singularities, or concentrations, and so on.

Of course one might also use some kind of measurement of sizes of subsets of M. This could entail diameters, volumes, or measurements of capacity. There are also many kinds of local measurements of oscillation or size that one can consider. As an extension of just taking suprema, one can take various local averages, or averages of powers of other quantities. Of course one can still bring in powers of the radius as before.

Even if one starts with measurements of localized behavior which are not so robust in the manner described before, one can transform them into more robust versions by taking localized suprema or averages or whatever afterwards. Frequently the kind of overall aggregations employed have this kind of robustness included in effect, and one can make some sort of rearrangement to put this in starker relief. Let us also note that one often has local measurements which can be quite different on their own, but in some overall aggregation lead to equivalent classes of functions and similar measurements of their behavior.

There are various moments, differences, and higher-order oscillations that can be interesting. As a basic version of this, one can consider oscillations of $f(x)$ in terms of deviations from something like a polynomial of fixed positive degree, rather than simply oscillations from being constant, as with $\mathrm{osc}(x, t)$. This can be measured in a number of ways.

However, for these kinds of higher-order oscillations, additional structure of the metric space is relevant. On Euclidean spaces, or subsets of Euclidean spaces, one can use ordinary polynomials, for instance. This carries over to the much-studied setting of nilpotent Lie groups equipped with a family of dilations, where one has polynomials as in the Euclidean case, with the degrees of the polynomials defined in a different way using the dilations.

These themes are closely related to having some kind of derivatives around. Just as there are various ways to measure the size of a function, one can get various measurements of oscillations looking at measurements of sizes of derivatives. It can also be interesting to have scales involved in a more active manner, and in any case there are numerous versions of ideas along these lines that one can consider.

3 Sub-Riemannian geometry

Let n be a positive integer, and consider $(2n + 1)$-dimensional Euclidean space, which we shall think of as

$$\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}. \tag{3.1}$$

The case where $n = 1$, so that this is basically just $\mathbf{R}^3$, is already quite interesting for us. We shall also be interested in

$$\mathbf{R}^n \times \mathbf{R}^n \tag{3.2}$$

and the obvious coordinate projection from the former to the latter given by

$$(x, y, s) \mapsto (x, y). \tag{3.3}$$

Define a smooth 1-form α on (3.1) by

$$\alpha = ds - \sum_{j=1}^{n} y_j \, dx_j, \tag{3.4}$$

using the usual coordinates (x, y, s) on (3.1). Thus if $p = (x, y, s)$ is a point in (3.1), then α at p, which we denote α_p, is a linear functional on the tangent space to (3.1) at p. Of course we can identify the tangent space to (3.1) at p with $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}$ itself in the usual way, and if $V = (v, w, u)$ is a tangent vector to (3.1) at p represented using this identification, then

$$\alpha_p(V) = u - \sum_{j=1}^{n} y_j \, v_j. \tag{3.5}$$

We can take the derivative $d\alpha$ of α in the sense of exterior differential calculus to get a 2-form on (3.1). Namely,

$$d\alpha = \sum_{j=1}^{n} dx_j \wedge dy_j. \tag{3.6}$$

If one takes the wedge product of α with n copies of $d\alpha$, then one gets

$$(n!) \, ds \wedge dx_1 \wedge dy_1 \cdots dx_n \wedge dy_n, \tag{3.7}$$

which is a nonzero $(2n+1)$-form on (3.1) that is n factorial times the standard volume form.

If $p = (x, y, s)$ is a point in (3.1), then we get a special linear subspace H_p of the tangent space to (3.1) at p which is the kernel of α_p. In other words, H_p consists of the tangent vectors $V = (v, w, u)$ such that

$$u - \sum_{j=1}^{n} y_j \, v_j = 0. \tag{3.8}$$

Thus H_p has dimension $2n$ at every point p.

If f is a smooth real-valued function on (3.1) which is never equal to 0, then $f \, \alpha$ is also a 1-form on (3.1) which vanishes exactly at the tangent vectors in H_p at a point p. By the usual rules of exterior differential calculus,

$$d(f \, \alpha) = df \wedge \alpha + f \, d\alpha. \tag{3.9}$$

One can check that the wedge product of $f\alpha$ with n copies of $d(f\alpha)$ is the same as f^{n+1} times the wedge product of α with n copies of $d\alpha$.

A nice feature of $d\alpha$ is that it can be viewed as the pull-back of a 2-form from (3.2). Basically this simply means that $d\alpha$ does not contain any ds's and the coefficients depend only on x, y. However, α contains ds as an important term, and α is not the pull-back of a form from (3.2).

For each point $p = (x, y, s)$ in (3.1), let F_p denote the 1-dimensional linear subspace of the tangent space to (3.1) at p consisting of vectors of the form $(0, 0, t)$, $t \in \mathbf{R}$. This is the same as the subspace of tangent vectors to the fiber of the mapping from (3.1) to (3.2) through p, which is also the same as the set of tangent vectors in the kernel of the aforementioned mapping. Notice that F_p and H_p are complementary subspaces of the tangent space to (3.1), which is to say that every vector $V = (v, w, u)$ in the tangent space at p can be written in a unique way as a sum of vectors in F_p and H_p, namely

$$V = \Big(0, 0, u - \sum_{j=1}^{n} y_j \, v_j\Big) + \Big(v, w, \sum_{j=1}^{n} y_j \, v_j\Big). \tag{3.10}$$

Let I be an interval in the real line with positive length, and which may be unbounded. Suppose that $\gamma(t)$ is a continuous function defined for t in I and with values in (3.1), so that $\gamma(t)$ defines a continuous path in (3.1). Let us assume that $\gamma(t)$ is continuously-differentiable on I, so that the derivative

$$\dot{\gamma}(t) = \frac{d}{dt}\gamma(t) \tag{3.11}$$

exists for each $t \in I$ and is continuous on I. More precisely, if t is in the interior of I, then the derivative is taken in the usual sense, while if t is an element of I which is also an endpoint of I, then one uses a one-sided derivative.

We say that this path $\gamma(t)$ is *horizontal* if

$$\dot{\gamma}(t) \in H_{\gamma(t)} \tag{3.12}$$

for each $t \in I$. This is equivalent to asking that

$$\alpha(\dot{\gamma}) = 0 \tag{3.13}$$

along the curve, which is to say that

$$\alpha_{\gamma(t)}(\dot{\gamma}(t)) = 0 \tag{3.14}$$

for all $t \in I$. If we write $\gamma(t)$ more explicitly as $(x(t), y(t), s(t))$, then this becomes

$$\dot{s}(t) = \sum_{j=1}^{n} y_j(t)\, \dot{x}_j(t) \tag{3.15}$$

for all $t \in I$.

Let $\gamma_0(t)$ be the projection of $\gamma(t)$ into (3.2), which means that $\gamma_0(t) = (x(t), y(t))$. If $\gamma(t)$ is a horizontal curve in (3.1), then $\gamma(t)$ is uniquely determined by its projection $\gamma_0(t)$ and the value of $\gamma(t)$ at a single point $t = t_0 \in I$. Indeed, if $\gamma(t) = (x(t), y(t), s(t))$ is horizontal, then $\dot{s}(t)$ can be expressed in terms of $\dot{x}(t)$ and $y(t)$ as before, and $s(t)$ is determined by this and the value of $s(t)$ at one point t_0.

Conversely, suppose that we are given a continuously-differentiable curve $\gamma_0(t)$ in (3.2) defined for $t \in I$. Also let t_0 be an element of I, and suppose that s_0 is some real number. Then there is a continuously-differentiable curve $\gamma(t) = (x(t), y(t), s(t))$, $t \in I$, in (3.1) whose projection into (3.2) is equal to $\gamma_0(t)$ and such that $s(t_0) = s_0$. Namely, one can compute what $\dot{s}(t)$ should be in terms of $\dot{x}(t)$ and $y(t)$, and integrate that using also $s(t_0) = s_0$ to get $s(t)$ for all $t \in I$.

It is not too difficult to show that each pair of points p, q in (3.1) can be connected by a continuously-differentiable curve which is horizontal. One can start by taking the prjections of these two points to get two points in (3.2) which can be connected by all sorts of curves. These curves can be lifted to horizontal curves in (3.1) which begin at p, as in the preceding paragraph.

In general a lifted path like this may not go to q, and one can check that there are plenty of choices of paths in (3.2) for which the lifting will have this property.

This leads to a very interesting kind of geometry in (3.1). Namely, one defines the distance between two points p, q to be the infimum of the lengths of the horizontal paths joining p to q. Of course the standard Euclidean metric on (3.1) can be defined by minimizing the lengths of all paths from p to q. The restriction to horizontal paths makes the distance increase, although one can check that the resulting metric is still compatible with the usual topology on (3.1).

4 Hyperbolic groups

Let Γ be a group, and let F be a finite set of elements of Γ. By a *word* over F we mean a formal product of elements of F and their inverses. Every word over F determines an element of the group Γ, simply using the group operations. The "empty word" is considered a word over F, which corresponds to the identity element of Γ.

If z is a word over F, then the length of z is denoted $L(z)$ and is the number of elements of F such that they or their inverses are used in z, counting multiplicities. A word z is said to be irreducible if it does not contain an $\alpha \in F$ next to α^{-1}, i.e., so that all obvious cancellations have been made. If a word z over F corresponds to the identity element of Γ, then z is said to be trivial.

A finite subset F of a group Γ is a set of *generators* of Γ if every element of Γ corresponds to a word over F. A group is said to be *finitely-generated* if it has a finite set of generators. Let us make the convention that a generating set F of a group Γ should not contain the identity element of Γ.

Suppose that Γ is a group and that F is a finite set of generators of Γ. The *Cayley graph* associated to Γ and F is the graph consisting of the elements of Γ as vertices with the provision that γ_1, γ_2 in Γ are adjacent if $\gamma_2 = \gamma_1\, \alpha$, where α is an element of F or its inverse. Thus this relation is symmetric in γ_1 and γ_2.

A finite sequence $\theta_0, \theta_2, \ldots, \theta_k$ of elements of Γ is said to define a *path* if θ_j, θ_{j+1} are adjacent in the Cayley graph for each j, $0 \le j \le k-1$. The *length* of this path is defined to be k. We include the degenerate case where $k = 0$, so that a single element of Γ is viewed as a path of length 0.

If ϕ, ψ are elements of Γ, then the *distance* between ϕ and ψ is defined to be the shortest length of a path that connects ϕ to ψ. In particular, note that for any two elements ϕ, ψ in Γ there is a path which starts at ϕ and ends at ψ. To see this, one can write ψ as $\phi\,\beta$ for some β in Γ, and then use the assumption that Γ is generated by F to obtain a path from ϕ to ψ one step at a time.

These definitions are invariant under *left* translations in Γ. In other words, if δ is any fixed element of Γ, then the tranformation $\gamma \mapsto \delta\,\gamma$ on Γ defines an automorphism of the Cayley graph, and it also preserves distances between elements of Γ. This follows from the definitions, since the Cayley graph was defined in terms of right-multiplication by generators and their inverses.

A basic fact is that this definition of distance does not depend too strongly on the choice of generating set F, in the sense that if one has another finite generating set, then the two distance functions associated to these generating sets are each bounded by a constant multiple of the other. This is not difficult to check, by expressing each generator in one set as a finite word over the other set of generators. There are only a finite number of these expressions, so that their maximal length is a finite number.

Let us continue with the assumption that we have a fixed generating set F for the group Γ. Suppose that R is a finite set of words over F. We say that R is a set of *relations* for Γ if every element of R is a trivial word. The inverses of elements of R are also then trivial words, as well as conjugates of elements of R. That is, if r is an element of R and u is any word over F, then $u\,r\,u^{-1}$ is the conjugate of r by u, and it is a trivial word since r is. Products of conjugates of elements of R and their inverses are trivial words too, as well as words obtained from these through cancellations, i.e., by cancelling $\alpha\,\alpha^{-1}$ and $\alpha^{-1}\,\alpha$ whenever α is an element of F. The combination of F and a set R of relations defines a *presentation* of Γ if every word over F which corresponds to the identity element of Γ can be obtained in this manner. The empty word is viewed as being equal to the empty product of relations, so that it is automatically included. A group Γ is said to be *finitely-presented* if there is a presentation with a finite set of generators and a finite set of relations. For instance, if Γ is the free group with generators in F, then one can take R to be the set consisting of the empty word, and this defines a presentation for Γ.

Let us call a word over F *trivial* if it corresponds to the identity element

of Γ. Suppose that w is a trivial word, with

$$w = \beta_1 \beta_2 \cdots \beta_n, \tag{4.1}$$

where each β_i is an element of F or an inverse of an element of F. This leads to a path $\theta_0, \theta_1, \ldots, \theta_n$, where θ_0 is the identity element of Γ and θ_j is equal to $\beta_1 \beta_2 \cdots \beta_j$ when $j \geq 1$. Because w is a trivial word, θ_n is also the identity element in Γ, which is to say that this path is a loop that begins and ends at the identity element.

Fix a finite set R of relations, so that F and R give a presentation for Γ. Let w be a trivial word over F which is also irreducible. Define $A(w)$ to be the smallest nonnegative integer A for which there exist relations $r_1, r_2, \ldots, r_k$ in R, integers $b_1, b_2, \ldots, b_k$, and words $u_1, u_2, \ldots, u_k$ over F such that the expression

$$u_1 r_1^{b_1} u_1^{-1} u_2 r_2^{b_2} u_2^{-1} \cdots u_k r_k^{b_k} u_k^{-1} \tag{4.2}$$

can be reduced to w after cancellations as before,

$$\sum_{j=1}^{k} L(u_j) \leq A, \tag{4.3}$$

and

$$\sum_{j=1}^{k} |b_j| \, L(r_j)^2 \leq A. \tag{4.4}$$

Here if z is a word over F and b is an integer, then z^b is defined in the obvious manner, by simply repeating z b times when $b \geq 0$, or repeating z^{-1} $-b$ times when $b < 0$. A representation of this type for w necessarily exists, since F and R give a presentation for Γ.

The group Γ is said to be *hyperbolic* if there is a nonnegative real number $C_0 \geq 0$ so that

$$A(w) \leq C_0 \, L(w) \tag{4.5}$$

for all irreducible trivial words w. The property of hyperbolicity does not depend on the choice of finite presentation for Γ, and in fact there are other definitions for which one only needs to assume that Γ is finitely generated, and the existence of a finite presentation is then a consequence. This characterization of hyperbolicity is discussed in Section 2.3 of [Gro1]. Some examples of hyperbolic groups are finitely-generated free groups and the fundamental groups of compact connected Riemannian manifolds without boundary

and strictly negative curvature. In particular, this includes the fundamental group of a closed Riemann surface with genus at least 2.

Let M be a nonempty set. A nonnegative real-valued function $d(x, y)$ on the Cartesian product $M \times M$ is said to be a *quasimetric* if $d(x, y) = 0$ exactly when $x = y$, $d(x, y) = d(y, x)$ for all $x, y \in M$, and

$$d(x, z) \le C\Big(d(x, y) + d(y, z)\Big) \tag{4.6}$$

for some positive real number C and all $x, y, z \in M$. If this last condition holds with $C = 1$, then $d(x, y)$ is said to be a *metric* on M.

If $d(x, y)$ is a quasimetric on M and a is a positive real number, then $d(x, y)^a$ is also a quasimetric on M. If $d(x, y)$ is a metric on M and a is a positive real number such that $a \le 1$, then $d(x, y)^a$ is a metric on M too. These statements are not difficult to verify. There is a very nice result going in the other direction, which states that if $d(x, y)$ is a quasimetric on M, then there are positive real numbers C', δ and a metric $\rho(x, y)$ on M such that

$$C'^{-1}\, \rho(x, y)^\delta \le d(x, y) \le C'\, \rho(x, y)^\delta \tag{4.7}$$

for all $x, y \in M$. See [MacS1].

If $d(x, y)$ is a quasimetric on M, then one has many of the same basic notions as for a metric, such as convergence of sequences, open and closed sets, dense subsets, and so on. For instance, it makes sense to say that M is separable with respect to a quasimetric if it has a subset which is at most countable and also dense, and one can define the topological dimension for M as in [HurW]. The diameter of a subset can be defined in the usual manner using the quasimetric, and this permits one to define the Hausdorff dimension of a nonempty subset of M. A famous result about metric spaces is that the topological dimension is always less than or equal to the Hausdorff dimension. See Chapter VII of [HurW]. This does not work for quasimetrics in general, and it cannot possibly work. For if $(M, d(x, y))$ is a quasimetric space with Hausdorff dimension s and a is a positive real number, then $(M, d(x, y)^a)$ has Hausdorff dimension s/a, while the topological dimension of $(M, d(x, y)^a)$ is the same as that of $(M, d(x, y))$.

Let Γ be a finitely-presented group which is hyperbolic. Associated to Γ is a space Σ which is a kind of "space at infinity" or ideal boundary of Γ, consisting of equivalence classes of asymptotic directions in Γ. This space is a compact Hausdorff topological space of finite dimension, as on p110-1 of [Gro1], and it contains a copy of the Cantor set as soon as it has at least three

elements. If Σ has at most two elements, then Γ is said to be *elementary.* For a free group with at least two generators the space at infinity is homeomorphic to a Cantor set, while $\mathbf{Z}$, a free group with one generator, is elementary and has two points in the space at infinity. If Γ is the fundamental group of a closed Riemann surface of genus at least 2, then Σ is homeomorphic to the unit circle in $\mathbf{R}^2$. More generally, if Γ is the fundamental group of a compact n-dimensional Riemannian manifold without boundary with strictly negative curvature, then Σ is homeomorphic to the unit sphere $\mathbf{S}^{n-1}$ in $\mathbf{R}^n$.

Actually, the space at infinity is defined for any hyperbolic metric space in [Gro1], and this can be specialized to a hyperbolic group. It is often preferable to work with metric spaces which are "geodesic", in the sense that any pair of points can be connected by a curve whose length is equal to the distance between the two points. It is often useful to think of a hyperbolic group as acting on a geodesic hyperbolic metric space by isometries, and to use that to study the space at infinity.

It does not customarily seem to be said this way, but I think it is fair to say that what are basically defined on the space at infinity are quasimetrics, at least initially. More precisely, it is more like the logarithm of a quasimetric, or, in other words, there is a one-parameter family of quasimetrics which are powers of each other. A few years ago Gromov casually asked about approximating quasimetrics by metric in the manner described before, and this is presumably the reason. In Section 7.2 of [Gro1] one takes a different route, in effect compactifying a geodesic hyperbolic metric space by looking at modified measurements of lengths of curves which take densities into account, densities that decay at infinity in a suitable manner.

In nice situations, such as hyperbolic groups, and universal coverings of compact Riemannian manifolds without boundary and strictly negative curvature in particular, there are doubling conditions on the space at infinity. Compare with [Pan1]. There are also interesting measures around, as in [Coo].

A well-known result of Borel [Bor1, Rag] says that simply-connected symmetric spaces can be realized as the universal covering of a compact manifold. If the symmetric space is of noncompact type and rank 1, it has negative curvature, and thus the fundamental group of the compact quotient, which is a uniform lattice in the group of isometries of the symmetric space, is a hyperbolic group. If the symmetric space is a classical hyperbolic space of dimension n, with constant negative curvature, then the space at infinity can be identified with a Euclidean sphere of dimension $n - 1$. If the symmetric

space is a complex hyperbolic space of complex dimension m, then the space at infinity can be identified topologically with a Euclidean sphere of real dimension $2m - 1$, but the geometry corresponds to a sub-Riemannian space when $m \geq 2$. For other symmetric spaces of noncompact type and rank 1, one again obtains topological spheres of dimension 1 less than the real dimension of the symmetric space, with more complicated sub-Riemannian structures.

5 p-Adic numbers

Let $\mathbf{Z}$ denote the integers, $\mathbf{Q}$ denote the rational numbers, and let $|\cdot|$ denote the usual absolute value function or modulus on the complex numbers $\mathbf{C}$. On the rational numbers there are other absolute value functions that one can consider. Namely, if p is a prime number, define the *p-adic absolute value function* $|\cdot|_p$ on $\mathbf{Q}$ by $|x|_p = 0$ when $x = 0$, $|x|_p = p^{-k}$ when $x = p^k m/n$, where k is an integer and m, n are nonzero integers which are not divisible by p. One can check that

$$|xy|_p = |x|_p \, |y|_p \tag{5.1}$$

and

$$|x + y|_p \leq |x|_p + |y|_p \tag{5.2}$$

for all $x, y \in \mathbf{Q}$, and in fact

$$|x + y|_p \leq \max(|x|_p, |y|_p) \tag{5.3}$$

for all $x, y \in \mathbf{Q}$.

Just as the usual absolute value function leads to the distance function $|x-y|$, the p-adic absolute value function leads to the p-adic distance function $|x - y|_p$ on $\mathbf{Q}$. With respect to this distance function, the rationals are not complete as a metric space, and one can complete the rationals to get a larger space $\mathbf{Q}_p$. This is analogous to obtaining the real numbers by completing the rationals with respect to the standard absolute value function. By standard reasoning the arithmetic operations and p-adic absolute value function extend from $\mathbf{Q}$ to $\mathbf{Q}_p$, with much the same properties as before. In this manner one gets the field of p-adic numbers. As a metric space, $\mathbf{Q}_p$ is complete by construction, and one can also show that closed and bounded subsets of $\mathbf{Q}_p$ are compact. This is also similar to the real numbers.

Note that the set $\mathbf{Z}$ of integers forms a bounded subset of $\mathbf{Q}_p$, in contrast to being an unbounded subset of $\mathbf{R}$. In fact, each integer has p-adic absolute value less than or equal to 1. There are general results about absolute value functions on fields to the effect that if the absolute values of integers are bounded, then they are less than or equal to 1, and the absolute value function satisfies the ultrametric version of the triangle inequality. See p28-9 of [Gou]. In this case the absolute value function is said to be *non-Archimedian*. If the absolute values of integers are not bounded, as in the case of the usual absolute value function, then the absolute value function is said to be *Archimedian*.

A related point is that the set $\mathbf{Z}$ of integers is a *discrete* subset of the real numbers. It has no limit points, and in fact the distance between two distinct integers is always at least 1. This is not the case in $\mathbf{Q}_p$, where $\mathbf{Z}$ is bounded, and hence precompact. Now consider $\mathbf{Z}[1/p]$, the set of rational numbers of the form $p^k n$, where k and n are integers. As a subset of $\mathbf{R}$, this is unbounded, and it also contains nontrivial sequences which converge to 0. Similarly, as a subset of $\mathbf{Q}_p$, it is unbounded and contains nontrivial sequences which converge to 0. As a subset of $\mathbf{Q}_l$ when $l \neq p$, $\mathbf{Z}[1/p]$ is bounded and hence precompact again. Using the diagonal mapping $x \mapsto (x, x)$, one can view $\mathbf{Z}[1/p]$ as a subset of the Cartesian product $\mathbf{R} \times \mathbf{Q}_p$. In this product, $\mathbf{Z}[1/p]$ is discrete again. Indeed, if $a = p^k b$ is a nonzero element of $\mathbf{Z}[1/p]$, where k, b are integers and b is not divisible by p, then either $|a|_p \geq 1$, or $|a|_p \leq 1$, in which case $k \geq 0$, and $|a| \geq 1$.

Similarly, $SL_n(\mathbf{Z})$, the group of $n \times n$ invertible matrices with entries in $\mathbf{Z}$ and determinant 1, is a discrete subgroup of $SL_n(\mathbf{R})$, the analogously-defined group of matrices with real entries. One can define $SL_n(\mathbf{Z}[1/p])$ and $SL_n(\mathbf{Q}_p)$ in the same manner, and using the diagonal embedding $x \mapsto (x, x)$ again, $SL_n(\mathbf{Z}[1/p])$ becomes a discrete subgroup of the Cartesian product $SL_n(\mathbf{R}) \times SL_n(\mathbf{Q}_p)$.

There are fancier versions of these things for making $\mathbf{Q}$ discrete, using "adèles", which involve p-adic numbers for all primes p. See [Wei].

Now let us turn to some aspects of analysis. With respect to addition, $\mathbf{Q}_p$ is a locally compact abelian group, and thus has a translation-invariant Haar measure, which is finite on compact sets, strictly positive on nonempty open sets, and unique up to multiplication by a positive real number. As in [Tai], there is a rich Fourier analysis for real or complex-valued functions on $\mathbf{Q}_p$, or $\mathbf{Q}_p^n$ when n is a positive integer.

Instead one can also be interested in $\mathbf{Q}_p$-valued functions on $\mathbf{Q}_p$, or on

a subset of $\mathbf{Q}_p$. It is especially interesting to consider functions defined by power series. As is commonly mentioned, a basic difference between $\mathbf{Q}_p$ and the real numbers is that an infinite series $\sum a_n$ converges if and only if the sequence of terms a_n tends to 0 as n tends to infinity. Indeed, the series converges if and only if the sequence of partial sums forms a Cauchy sequence, and this implies that the terms tend to 0, just as in the case of real or complex numbers. For p-adic numbers, however, one can use the ultrametric version of the triangle inequality to check that the partial sums form a Cauchy sequence when the terms tend to 0. In particular, a power series $\sum a_n \, x^n$ converges for some particular x if and only if the sequence of terms $a_n \, x_n$ tends to 0, which is to say that $|a_n|_p \, |x|_p^n$ tends to 0 as a sequence of real numbers.

Suppose that $\sum_{n=0}^{\infty} a_n \, x^n$ is a power series that converges for all x in $\mathbf{Q}_p$, which is equivalent to saying that $|a_n|_p r^n$ converges to 0 as a sequence of real numbers for all $r > 0$. Thus we get a function $f(x)$ defined on all of $\mathbf{Q}_p$, and we would like to make an analogy with entire holomorphic functions of a single complex variable. This is somewhat like the situation of starting with a power series that converges on all of $\mathbf{R}$, and deciding to interpret it as a function on the complex numbers instead.

In fact, let us consider the simpler case of a power series with only finitely many nonzero terms, which is to say a polynomial. As in the case of complex numbers, it would be nice to be able to factor polynomials. The p-adic numbers $\mathbf{Q}_p$ are not algebraically closed, and so in order to factor polynomials one can first pass to an algebraic closure. It turns out that the p-adic absolute value can be extended to the algebraic closure, while keeping the basic properties of the absolute value. See [Cas, Gou]. The algebraic closure is not complete in the sense of metric spaces with respect to the extended absolute value function, and one can take a metric completion to get a larger field to which the absolute values can be extended again. A basic result is that this metric completion is algebraically closed, so that one can stop here. Let us write $\mathbf{C}_p$ for this new field, which is algebraically closed and metrically complete.

Once one goes to the algebraic closure, one can factor polynomials. On $\mathbf{C}_p$ one has this property and also one can work with power series. In particular, since the power series $\sum_{n=0}^{\infty} a_n \, x^n$ converges on all of $\mathbf{Q}_p$, it also converges on all of $\mathbf{C}_p$, so that $f(x)$ can be extended in a natural way to $\mathbf{C}_p$. For that matter, one can start with a power series that converges on all of $\mathbf{C}_p$, where the coefficients are allowed to be in $\mathbf{C}_p$, and not just $\mathbf{Q}_p$.

Under these conditions, the function f can be written as a product of an element of $\mathbf{C}_p$, factors which are equal to x, and factors of the form $(1-\lambda_j\, x)$, where the λ_j's are nonzero elements of $\mathbf{C}_p$. In other words, the factors of x correspond to a zero of some order at the origin, while the factors of the form $(1-\lambda_j\, x)$ correspond to zeros at the reciprocals of the λ_j's. If $f(x)$ is a polynomial, then there are only finitely many factors, and this statement is the same as saying that $\mathbf{C}_p$ is algebraically closed. In general, each zero of $f(x)$ is of finite order, and there are only finitely many zeros within any ball of finite radius in $\mathbf{C}_p$. Thus the set of zeros is at most countable, and this condition permits one to show that the product of the factors mentioned above converges when there are infinitely many factors.

This representation theorem can be found on p113 of [Cas] and on p209 of [Gou]. It is analogous to classical results about entire holomorphic functions of a complex variable, with some simplifications. In the complex case, it is necessary to make assumptions about the growth of an entire function for many results, and the basic factors often need to be more complicated in order to have convergence of the product. See [Ahl, Vee] concerning entire holomorphic functions of a complex variable.

References

[Ahl] L. Ahlfors, *Complex Analysis*, third edition, McGraw-Hill, 1979.

[AmbS] L. Ambrosio and F. Serra-Cassano, eds., *Lecture Notes on Analysis on Metric Spaces*, Scuola Normale Superiore, Pisa, 2000.

[AmbT] L. Ambrosio and P. Tilli, *Selected Topics on "Analysis in Metric Spaces"*, Scuola Normale Superiore, Pisa, 2000.

[AusCG] P. Auscher, T. Coulhon, and A. Grigoryan, eds., *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, Contemporary Mathematics **338**, American Mathematical Society, 2003.

[BelR] A. Bellaïche and J.-J. Risler, editors, *Sub-Riemannian Geometry*, Birkhäuser, 1996.

[Bor1] A. Borel, *Compact Clifford–Klein forms of symmetric spaces*, Topology **2** (1963), 111–122.

[Bor2] A. Borel, *Semisimple Groups and Riemannian Symmetric Spaces*, Hindustan Book Agency, 1998.

[Bro1] K. Brown, *Buildings*, Springer Monographs in Mathematics, Springer-Verlag, 1989.

[Bro2] K. Brown, *What is a building?*, Notices of the American Mathematical Society **49** (2002), 1244-1245.

[Cas] J. Cassels, *Local Fields*, London Mathematical Society Student Texts **3**, Cambridge University Press, 1986.

[CoiW1] R. Coifman and G. Weiss, *Analyse Harmonique Non-Commutative sur Certains Espaces Homogènes*, Lecture Notes in Mathematics **242**, Springer-Verlag, 1971.

[CoiW2] R. Coifman and G. Weiss, *Extensions of Hardy spaces and their use in analysis*, Bulletin of the American Mathematical Society **83** (1977), 569–645.

[Coo] M. Coornaert, *Mesures de Patterson–Sullivan sur le bord d'un espace hyperbolique au sens de Gromov*, Pacific Journal of Mathematics **159** (1993), 241–270.

[CooDP] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et Théorie des Groupes: Les Groupes Hyperboliques de Gromov*, Lecture Notes in Mathematics **1441**, Springer-Verlag, 1990.

[CooP] M. Coornaert and A. Papadopoulos, *Symbolic Dynamics and Hyperbolic Groups*, Lecture Notes in Mathematics **1539**, Springer-Verlag, 1993.

[Eps+] D. Epstein, J. Cannon, D. Holt, S. Levy, M. Paterson, W. Thurston, *Word Processing in Groups*, Jones and Bartlett, 1992.

[Far] B. Farb, *Automatic groups: A guided tour*, L'Enseignement Mathématique (2) **38** (1992), 291–313.

[GhyH] É. Ghys and P. de la Harpe, editors, *Sur les Groupes Hyperboliques d'après Mikhael Gromov*, Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988, Birkhäuser, 1990.

[Gou] F. Gouvêa, *p-Adic Numbers: An Introduction*, Universitext, Springer-Verlag, 1993.

[Gro1] M. Gromov, *Hyperbolic groups*, in *Essays in Group Theory*, S. Gersten, editor, 75–263, Mathematical Sciences Research Institute Publications **8**, Springer-Verlag, 1987.

[Gro2] M. Gromov, *Asymptotic invariants of infinite groups*, Volume 2 of *Geometric Group Theory*, G. Niblo and M. Roller, editors, London Mathematical Society Lecture Note Series **182**, Cambridge University Press, 1993.

[GroP] M. Gromov and P. Pansu, *Rigidity of lattices: An introduction*, in *Geometric Topology: Recent Developments (Montecatini Terme, 1990)*, 39–137, Lecture Notes in Mathematics **1504**, Springer-Verlag, 1991.

[Gro+] M. Gromov et al., *Metric Structures for Riemannian and Non-Riemannain Spaces*, Birkhäuser, 1999.

[Har] P. de la Harpe, *Topics in Geometric Group Theory*, University of Chicago Press, 2000.

[Hei] J. Heinonen, *Lectures on Analysis on Metric Spaces*, Springer-Verlag, 2001.

[HurW] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1941.

[Jou] J.-L. Journé, *Calderón–Zygmund Operators, Pseudo-Differential Operators and the Cauchy Integral of Calderón*, Lecture Notes in Mathematics **994**, Springer-Verlag, 1983.

[Kig] J. Kigami, *Analysis on Fractals*, Cambridge University Press, 2001.

[Kra] S. Krantz, *A Panorama of Harmonic Analysis*, Carus Mathematical Monographs **27**, Mathematical Association of America, 1999.

[MacS1] R. Macías and C. Segovia, *Lipschitz functions on spaces of homogeneous type*, Advances in Mathematics **33** (1979), 257–270.

[MacS2] R. Macías and C. Segovia, *A decomposition into atoms of distributions on spaces of homogeneous type*, Advances in Mathematics **33** (1979), 271–309.

[Mon] R. Montgomery, *A Tour of Subriemannian Geometry, Their Geodesics and Applications*, American Mathematical Society, 2002.

[Ohs] K. Ohshika, *Discrete Groups*, Translations of Mathematical Monographs **207**, American Mathematical Society, 2002.

[Pan1] P. Pansu, *Dimension conforme et sphère à l'infini des variétés à courbure négative*, Annales Academiae Scientiarum Fennicae Ser. A I Math. **14** (1989), 177–212.

[Pan2] P. Pansu, *Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un*, Annals of Mathematics (2) **129** (1989), 1–60.

[PicW] M. Picardello and W. Woess, eds., *Random Walks and Discrete Potential Theory*, Cambridge University Press, 1999.

[Rag] M. Raghunathan, *Discrete Subgroups of Lie Groups*, Springer-Verlag, 1972.

[Rud] W. Rudin, *Principles of Mathematical Analysis*, third edition, McGraw-Hill, 1976.

[Sem] S. Semmes, *An introduction to analysis on metric spaces* and *An introduction to Heisenberg groups in analysis and geometry*, Notices of the American Mathematical Society **50** (2003), 438–443 and 640–646.

[Ser] J.-P. Serre, *Cohomologie des groupes discrets*, in *Propsects in Mathematics*, Annals of Mathematics Studies **70**, 77–169, Princeton University Press, 1971.

[Ste1] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[Ste2] E. Stein, *Harmonic Analysis: Real-Variable methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, 1993.

[SteW] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.

[Str] R. Strichartz, *Analysis on fractals*, Notices of the American Mathematical Society **46** (1999), 1199–1208.

[Tai] M. Taibleson, *Fourier Analysis on Local Fields*, Mathematical Notes **15**, Princeton University Press, 1975.

[VarSC] N. Varopoulos, L. Saloff-Coste, and T. Coulhon, *Analysis and Geometry on Groups*, Cambridge University Press, 1992.

[Vee] W. Veech, *A Second Course in Complex Analysis*, Benjamin, 1967.

[Wei] A. Weil, *Basic Number Theory*, second edition, Springer-Verlag, 1973.

[Wo] W. Woess, *Random Walks on Infinite Graphs and Groups*, Cambridge University Press, 2000.